R E P O R T   R E S U M E S

ED 018 053   24   EC 002 052

THE COMPREHENSION OF RAPID SPEECH BY THE BLIND, PART III.

BY- FOULKE, EMERSON
LOUISVILLE UNIV., KY.

REPORT NUMBER CRP-2430   PUB DATE 15 DEC 67
REPORT NUMBER BR-5-1050
GRANT OEG-4-10-127
EDRS PRICE MF-$0.25 HC-$1.24 29P.

DESCRIPTORS- *EXCEPTIONAL CHILD RESEARCH, *VISUALLY HANDICAPPED, *COMMUNICATION (THOUGHT TRANSFER), RESEARCH REVIEWS (PUBLICATIONS), SPEECH COMPRESSION, COMPREHENSION METHODS, LISTENING COMPREHENSION, LISTENING, AUDITORY PERCEPTION, COGNITIVE PROCESSES, AURAL STIMULI, SKILL DEVELOPMENT, COMMUNICATION SKILLS, LISTENING SKILLS, EVALUATION, BLIND,

A REVIEW OF THE RESEARCH ON THE COMPREHENSION OF RAPID SPEECH BY THE BLIND IDENTIFIES FIVE METHODS OF SPEECH COMPRESSION--SPEECH CHANGING, ELECTROMECHANICAL SAMPLING, COMPUTER SAMPLING, SPEECH SYNTHESIS, AND FREQUENCY DIVIDING WITH THE HARMONIC COMPRESSOR. THE SPEECH CHANGING AND ELECTROMECHANICAL SAMPLING METHODS AND THE NECESSARY APPARATUS HAVE BEEN DEVELOPED SUFFICIENTLY FOR ADEQUATE EVALUATION. TWO GENERAL APPROACHES TO EVALUATION OF TIME COMPRESSED SPEECH ARE TESTS OF ABILITY TO REPEAT MESSAGES ACCURATELY AND TESTS OF COMPREHENSION OF LISTENING SELECTIONS. FACTORS SHOWN TO AFFECT INTELLIGIBILITY AND COMPREHENSION INCLUDE STIMULUS VARIABLES (THE CONTEXT IN WHICH THE SPEECH SIGNAL IS PRESENTED AS WELL AS THE CHARACTERISTICS OF THE SIGNAL ITSELF) AND ORGANISMIC VARIABLES (THE LISTENER'S AGE, SEX, EDUCATION, INTELLIGENCE, AND PRIOR RELEVANT EXPERIENCE). RESULTS OF SEVERAL STUDIES, WHEN CONSIDERED COLLECTIVELY, INDICATE THAT WORK INTELLIGIBILITY IS HAMPERED WHEN WORDS ARE COMPRESSED BY THE SPEECH CHANGING METHOD. WHEN THE SAMPLING METHOD IS USED, A COMPRESSION OF CONSIDERABLE MAGNITUDE IS REQUIRED BEFORE WORK INTELLIGIBILITY IS SUBSTANTIALLY IMPAIRED. PERCEPTUAL AND COGNITIVE PROBLEMS CONFRONTING THE LISTENER OF COMPRESSED SPEECH ARE DISCUSSED. A MODEL FOR SHORT TERM MEMORY IS USED TO ACCOUNT FOR THE DIFFERENCES IN THE EFFECT ON WORD INTELLIGIBILITY AND LISTENING COMPREHENSION OFINCREASING WORD RATE. GRAPHS PRESENT RESULTS, AND A BIBLIOGRAPHY LISTS 41 ITEMS. (KH)

Dupl

5-1050-24
MUST

ED018053

INTERIM PROGRESS REPORT
Project No. 2430
Grant No. OE-4-10-127

THE COMPREHENSION OF RAPID SPEECH BY THE BLIND: PART III

December 15, 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

EC 002 052
ERIC

THE COMPREHENSION OF RAPID SPEECH BY THE BLIND: PART III

Project No. 2430
Grant No. OE-4-10-127

Emerson Foulke, Ph.D.

December 15, 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

University of Louisville
Louisville, Kentucky

THE PERCEPTION OF TIME COMPRESSED SPEECH [1]

Introduction

Time compressed speech is speech that is reproduced in less time than the time required for its original production. One effect of time compression is the acceleration of the rate at which speech sounds occur. If speech, when accelerated, remains comprehensible, the savings in listening time should be an important consideration in those situations in which extensive reliance is placed on aural communication. However, current data suggest that although individual words and short phrases may remain quite intelligible after considerable compression by the right method, when such words are combined to form meaningful sequences that exceed in length the immediate memory span for heard words, as in the case of a listening selection, comprehension begins to deteriorate at a much lower compression. Compression may affect the intelligibility of single words and the comprehension of connected discourse differently because different cognitive processes underlie the behavior on which their measurement is based. In the pages that follow, several methods for the time compression of speech will be described, and compared with respect to the way in which they influence the intelligibility of speech sounds and the comprehension of connected discourse. Following this, an effort will be made to analyze the perceptual and cognitive problems confronting the listener to compressed speech.

Alternative Methods for the Time Compression of Speech

The Speed Changing Method

The word rate of recorded speech may be changed simply by reproducing it at a different tape or record speed than the speed used in recording the original speech production. If the playback speed is slower than the recording speed, the rate at which speech sounds occur is reduced, and the speech is expanded in time. If the playback speed is increased, the rate at which speech sounds occur is increased, and speech is compressed in time. Changing speech rate in this manner is, technically speaking, quite easy. However, the change in rate is accompanied by a shift in the frequency components of the voice signal that is proportional to the change in tape or record speed. For instance, if the speed is doubled, the component frequencies of a voice signal will be doubled, and the overall vocal pitch will be raised one octave. As will be shown later, this distortion in vocal pitch interferes with the intelligibility of time compressed spoken words.

[1] The author's research, reviewed in this paper, was performed at the University of Louisville, with financial support from the Office of Education under projects 1005, 1370, and 2430.

Sampling Methods

In 1950, Miller and Licklider demonstrated the signal redundancy in spoken words by deleting segments of the speech signal. This was accomplished by a switching arrangement that permitted a recorded speech signal to be turned off periodically, during its reproduction. They found that as long as these interruptions occurred at a frequency of ten times per second or more, the interrupted speech was easily understood. The intelligibility of monosyllabic words did not drop below 90%, until 50% of the speech signal had been discarded. Thus, it appeared that a large portion of the speech signal could be discarded without a serious disruption of communication. Garvey (1953a), taking cognizance of these results, reasoned that if the samples of a speech signal remaining after periodic interruption could be abutted in time, the result should be time compressed intelligible speech without distortion in vocal pitch. To test this notion, he prepared a tape on which speech had been recorded, by periodically cutting out short segments of tape, and by splicing the free ends of the retained tape together again. Reproduction of this tape achieved the desired effect. Garvey's method was, of course, too cumbersome for any but research purposes. However, the success of the general approach having been shown, an efficient technique for accomplishing it was not long to follow.

In 1954, Fairbanks, Everitt, and Jaeger published a description of an electromechanical apparatus which makes possible the time compressed or expanded reproduction of tape recorded speech. In the Fairbanks apparatus, a continuous tape loop passes over a recording head used to place on the tape the signal that is to be compressed. Next, it passes over the device used to reproduce samples of this signal. Finally, it passes over an erase head that removes the signal from the tape loop so that the tape can be rerecorded on the next cycle. The sampling device is a cylinder with four playback heads embedded in it and equally spaced around its circumference. The tape, in passing over the curved surface of this cylinder, makes contact with approximately one quarter of its circumference. When the cylinder is stationary, and one of the playback heads is contacted by the moving tape, the signal on the tape is reproduced as recorded. However, when the apparatus is adjusted for some amount of compression, the cylinder bearing the four playback heads begins to rotate in the direction of tape motion. Under these conditions, each of the four heads in turn makes and then loses contact with the tape. Each head reproduces the material on the portion of the tape with which it makes contact. When, as it rotates, the cylinder has arrived at a position at which one head is just losing contact with the tape while the preceding head is just making contact with the tape, the segment of tape that is wrapped around the cylinder between these two heads never makes contact with a reproducing head and is therefore not reproduced. The segment of the tape that is eliminated from the reproduction in this manner is always the same length, one quarter of the circumference of the cylinder. The amount of speech compression depends upon the frequency with which

these tape segments are eliminated, and this frequency depends, in turn, upon the rotational speed of the cylinder. The temporal value of the segments of tape that are not reproduced depends upon the rate of motion of the tape loop, since this determines the amount of tape that will pass over the recording head during a given time interval. Since the cylinder rotates in the direction of tape motion, the speed of the tape loop in relation to the surface of the cylinder is reduced, with the result that the retained samples scanned by the four playback heads are reproduced at a lower frequency. The output of the compressor is recorded on tape, and this tape is reproduced at a speed that is enough faster than the speed at which it was recorded to restore the frequencies represented in the signal to their original value. The final product of this process is time compressed speech that is unaffected with respect to vocal pitch.

Mr. Anton Springer, relying upon the same basic principles, developed a device in which the mode of operation was modified for greater convenience.[2] In the Springer approach, the continuous tape loop, the record head and the erase head are eliminated. Previously recorded tape passes from a supply reel, over the surface of the cylinder bearing the four playback heads, and then to a takeup reel. The tape is sampled in the manner just described. However, as the cylinder rotates in the direction of tape motion, the speed of the tape is increased by an amount sufficient to hold tape speed constant, in relation to the surface of the cylinder over which it passes. Thus, the output of the Springer device is already compressed in time, and without distortion in vocal pitch.

A computer may also be used to compress speech by the sampling method (Scott, 1965). In this approach, speech that has been transduced to electrical form (for example, the output of a microphone or tape reproducing head) is temporally segmented by an analog-to-digital converter and these segments are stored in the computer. Then, the computer samples these segments according to a sampling rule for which it has been programmed (for example, discard every third segment). The duration of both retained and discarded segments can be varied over a wide range. The retained samples are abutted in time and fed to the input of a digital-to-analog converter, and the signal at the output of this converter, compressed in time, is appropriate for transduction to acoustical form again.

Electromechanical compressors, of the Fairbanks or Springer type, are unselective with respect to the parts of a message that are discarded. Portions are discarded on a periodic basis, and may be discarded anywhere within or between words. It is quite unlikely that a given message would be sampled in exactly the same way on consecutive passes through such a device. With the computer, it is feasible to employ a great variety of sampling rules. For instance, the computer might be programmed to dispose of empty time intervals between words, and to sample the time intervals occupied by words differently, discarding larger fractions of those speech sounds with higher signal redundancy.

---

[2]The current version of the Springer device, known as the Information Rate Changer, is distributed in this country by Gotham Audio Corporation, 2 West 46th Street, New York, New York 10036.

From what has just been said, it would appear that the computer, because of its flexibility, offers the most satisfactory method for the time compression of speech. This may ultimately prove to be the case. However, at present, computer time is too expensive to justify the employment of a computer in this capacity for any but research purposes.

The technique of speech synthesis suggests another possibility for the production of accelerated speech without distortion in vocal pitch (Campanella, 1967). The speech synthesizer includes a group of acoustical generators with outputs that approximate the sounds represented in normal human speech, and a program of instructions for activating these generators for the proper durations, at the proper intensities, and in the proper sequence. These instructions may be varied to produce speech at any desired rate. Though this method has, as yet, received little development, it should share with the computer the ability to shorten speech sounds in accordance with their signal redundancy.

Another device for the time compression of speech, now under construction at the American Foundation for the Blind, is the harmonic compressor, an outgrowth of research at Bell Laboratories. In this approach, a speech signal is passed through an elaborate filter network that divides the speech spectrum into a large number of narrow frequency bands. The portion of the signal appearing in each of these bands is then reduced in frequency by one half and these derived signals are combined again to produce speech, the frequency of which has been reduced by one half. If a recording of this speech is reproduced at twice the recording speed, the result is speech that has been compressed to 50% of the original production time without distortion in vocal pitch. Since the prototype of the harmonic compressor is still under construction, there has, as yet, been no opportunity to evaluate its output. A limitation of the harmonic compressor that may prove to be serious is that it cannot be adjusted for any desired amount of compression. It can only shorten the time required for the reproduction of a message by one half.

To summarize, five methods for the acceleration of speech have been discussed. Three of these -- the computer sampling method, the speech synthesis method, and the frequency dividing method employed by the harmonic compressor -- have not yet been developed enough to permit an adequate evaluation of their products. For the remaining two methods -- the speed changing method and the electromechanical sampling method -- the necessary apparatus is readily available and of good quality. By now, a number of experiments evaluating the product of these methods, have been performed, and it is possible to draw some conclusions regarding their relative merits.

The Evaluation of Time Compressed Speech

Two general approaches have been employed in the evaluation of time compressed speech -- tests of the ability to repeat brief messages accurately, and tests of the comprehension of listening selections. Brief message reproduction is taken as an index of the intelligibility of

time compressed speech. A procedure typical of this approach is one in which single words are compressed in time by some amount, and presented, one at a time, to a listener. The listener's task is to reproduce these words, orally or in writing, and his intelligibility score is the fraction, usually expressed as a percent, of correctly identified words. This procedure is sometimes referred to as an articulation test (Miller, 1954, p. 60).

Disjunctive RT (reaction time) may also be taken as an index of intelligibility (Foulke, 1965). The underlying rationale in this case is that in the disjunctive RT experiment, reduced intelligibity means reduced discriminability. It has been shown that as stimuli are made less discriminable, choice RT is increased (Woodworth & Schlosberg, 1954, p. 33). The procedure, under this approach, is to acquaint S with a list of words for example three, and then to present them to him one at a time in random order for identification. Subject indicates his choice with a discriminative response, for instance, pressing the appropriate one of several response keys. Subject can then be scored for RT and accuracy. The experiment is performed with words that have been compressed in time by several amounts, and changes in RT and/or accuracy are regarded as indicative of changes in intelligibility. The RT method may be more sensitive than other methods, since a change in the amount of compression may produce a change in RT to words that are discriminated without error.

Calearo and Lazzaroni (1957) report the use of a procedure familiar to those in clinical audiology for detecting the effect of compression. The minimum intensity required for words to be intelligible is determined for words at several levels of compression. Threshold is determined for words at several levels of compression. Threshold intensity is defined as that intensity at which some percent of a list of words, for example 50%, are correctly identified. If a change in the compression in a list of words is accompanied by a change in threshold intensity for that list, it is concluded that time compression has altered intelligibility.

The other common approach in evaluating the effects of the acceleration of speech is one in which the listener first hears a listening selection at some accelerated word rate, and then is tested for knowledge of the facts and implications of that selection. Any kind of test may be used, but researchers have, in general, preferred objective tests with specifiable reliability.

Factors Affecting the Intelligibility and Comprehension of Time Compressed Speech

Factors that have been shown to have an effect upon the intelligibility and the comprehension of time compressed speech can be divided into two general classes. One class includes stimulus variables associated with the context in which the speech signal is presented, and characteristics of the signal itself. The second class includes organismic

variables, such as the listener's age, sex, education, intelligence, and prior relevant experience. No effort will be made here to present the results of studies of all of these factors. Instead, only that research will be reviewed which concerns factors believed to have an important bearing on the argument to be developed subsequently regarding differences in the mediating processes that underlie the demonstration of intelligibility and comprehension. Those readers interested in a more complete review of research regarding such factors are referred to "A Review of Research on Time Compressed Speech" (Foulke & Sticht, 1967a).

Intelligibility and the Method of Compression

The intelligibility of time compressed words depends upon the method used for compression. When the speed changing method is used, reproduction of words in two-thirds of the original production time results in a loss of intelligibility of 40% or more (Fletcher, 1929; Garvey, 1953a; Klumpp & Webster, 1961). On the other hand, Garvey (1953a) found only a 10% loss in the intelligibility of words that were reproduced in 40% of the original production time by means of his manual sampling method, and a 50% loss in the intelligibility of words reproduced in 25% of the original production time. Kurtzrock (1957), using the electromechanical sampling method, obtained an intelligibility score of 50% for a list of words reproduced in only 15% of the original production time. Using a similar method and similar materials, Fairbanks and Kodman (1957) obtained an intelligibility score of 57% for words reproduced in 13% of the original production time.

The relationships displayed in Fig. 1 (based on the figure in Garvey's 1953a article) are fairly typical. In this figure, the percent of the original production time required for the compressed reproduction of words is shown on the X-axis. Mean percent of words correctly identified is shown on the Y-axis. The dotted curve describes results obtained with the speed changing method, and the solid curve describes results obtained with the manual sampling method.

Compression by either method increases the rate at which the discriminable elements of speech occur. However, whereas vocal pitch is unaffected by the sampling method, it is elevated by the speed changing method. The difference in the intelligibility of words compressed by the two methods is probably due to the distortion in vocal pitch, since this is the factor that is not common to the two methods.

Intelligibility and the Sampling Rule

The sampling period of speech that is to be compressed in time by the sampling method is the interval between the onsets of consecutive retained portions of the message. Compression is accomplished by discarding part of this interval. It is the ratio of the retained to the

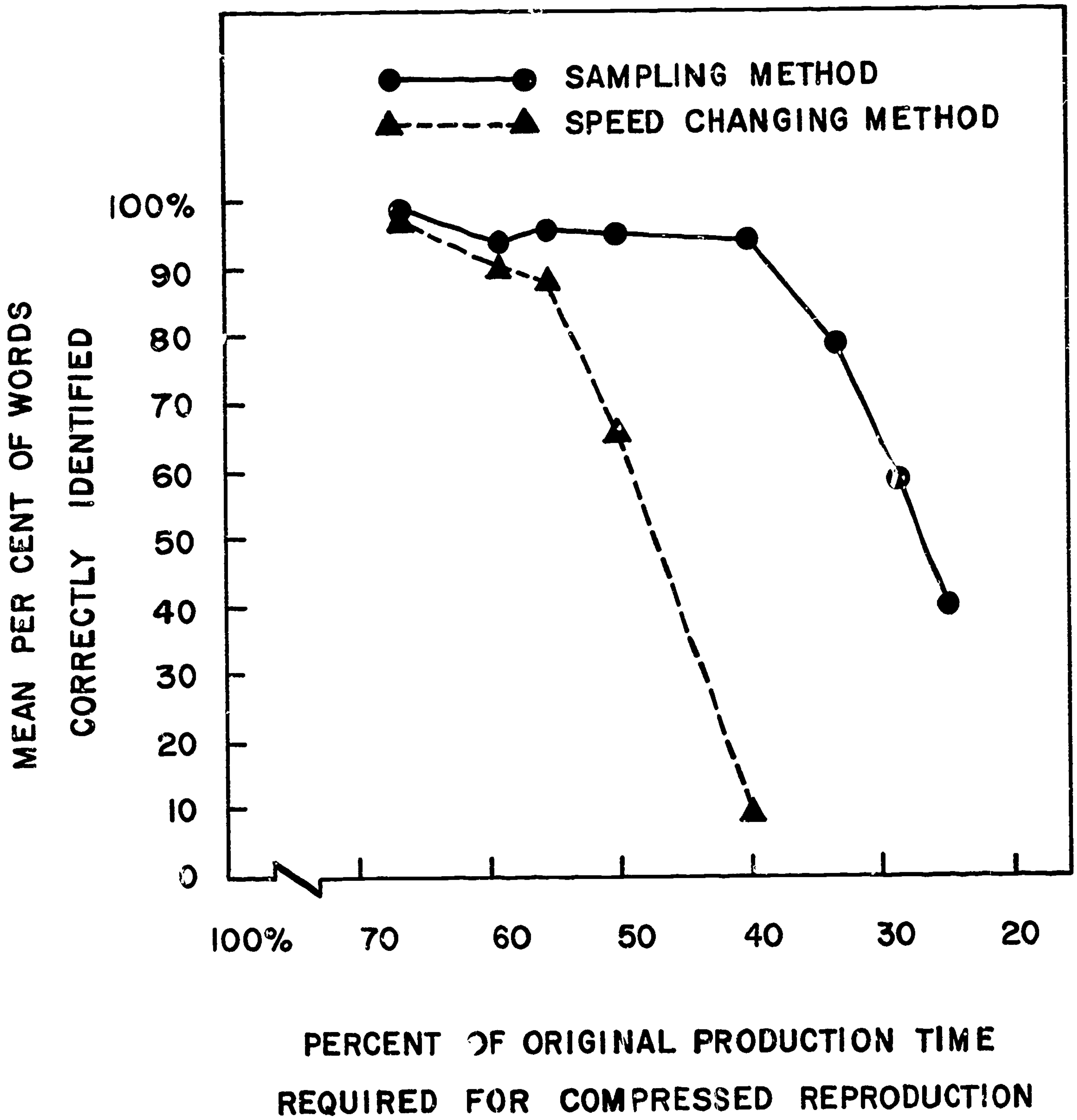

Fig. 1. Word intelligibility as a function of method of compression.

discarded portions of sampling periods that determines the amount of compression. If ten msec. (milliseconds) of a twenty msec. sampling period, or thirty msec. of a sixty msec. sampling period are retained, the result is the same -- 50% compression. For any given sampling period, changing the ratio of retained to discarded portions changes the amount of compression. When the sampling method is used, the effect that a given amount of compression will have on the intelligibility of words depends upon the duration of the discarded portion of the sampling period, and hence upon the duration of the sampling period itself. The duration of the discarded portion of the sampling period must be short relative to the durations of the speech sounds to be sampled. If it is not, a speech sound may fall entirely within the discarded portion of a sampling period, in which case it would not be sampled at all.

With spondaic words, compressed to 50% of their original durations, Garvey (1953a), using discard intervals of 40 msec., 60 msec., 80 msec., and 100 msec., found corresponding intelligibility scores of 95.33%, 95.67%, 95%, and 85.67%. In a two factor experiment in which five discard intervals were employed for each of eight compressions, Fairbanks and Kodman (1957) also found a substantial loss in intelligibility when the duration of the discard interval exceeded 80 msec. This was true at all eight compressions.

The intelligibility of a word may be degraded if the word is sampled too frequently. Speech that is compressed in time by the sampling method consists of a succession of abutted samples of the original speech if the transitions from sample to sample in this succession occur with sufficient frequency, the result is an audible tone with definite pitch. If the sampling rate is high enough, the pitch of this tone will intrude into the speech spectrum, and mask some speech frequencies. Fairbanks and Kodman (1957), using a discard interval of 10 msec., found 90% intelligibility for words compressed to 20% of their original duration. When this discard interval was changed to 40 msec., they found 94% intelligibility. When a 10 msec. discard interval is used to compress speech to 40% of its original time, the retained samples are 2.5 msec. in duration, and they occur at a rate of 400 per second. The 400 cycle tone corresponding to this rate is well within the speech spectrum and might be expected to interferewith intelligibility. If, on the other hand, a 40 msec. discard interval is used in compressing speech 20% of its original duration, the retained samples are 10 msec. in length, and they occur at a rate of 100 per second. The audible tone of corresponding frequency is below the speech spectrum in this case, and there should be little interference.[3]

Intelligibility and the Rate of Occurance of Speech Sounds

Garvey (1953a) compared the intelligibility of words compressed in time by the sampling method with the intelligibility reported by Miller

---

[3]The equipment used for the compressed reproduction of speech in the experiments reported hereafter. samples with a discard interval of 40 msec.

and Licklider (1950) for words that had been interrupted periodically. Garvey's words and Miller and Licklider's words were treated alike in that portions of sampling periods were discarded. However, the retained samples of Garvey's words were abutted in time to produce time compressed speech, while the retained samples of Miller and Licklider's words were not abutted and the resulting speech, though interrupted, was not compressed in time. There was no difference between the intelligibility of time compressed words and interrupted words when 50% of each word was discarded. However, when 62% of each word was discarded, interrupted words were 40% more intelligible than time compressed words. Since the two groups of words were alike with respect to the amount of speech information that had been discarded, the poorer intelligibility of the time compressed words, when 62% of the speech information had been discarded, was probably due to the accelerated rate of occurence of speech sounds.

Intelligibility and Word Structure

In the study in which the number of phonemes in a word was varied from three to nine, Henry (1966) found that increasing the number of phonemes imporved the intelligibility of words that had been compressed by the sampling method. In a similar vein, Klumpp and Webster (1951) found short phrases, compressed in time by the speed changing method, to be more intelligible than single words. The findings of Henry and of Klumpp and Webster are probably explained by the cues S's can derive from the context of multiphonemic words and short phrases.

Intelligibility and Prior Experience of the Listener

Using the sampling method, Fairbanks and Kodman (1957) compressed words enough so that they were comparable to the interrupted words of Miller and Licklider with respect to the amount of speech information discarded. They found the compressed words to be more intelligible than the interrupted words. However, the S's of Fairbanks and Kodman had received extensive familiarization with the words to be identified, before the tests were made, whereas the S's of Miller and Licklider were relatively naive. Miller and Licklider (1950), using interrupted words, and Garvey (1953b), using words compressed in time by the manual sampling method, found that repeated exposure to such words improved their intelligibility.

If a group of listeners agree that a particular speech sound in a word that has been compressed by the sampling method is unrecognizable, it may fairly be concluded that the difficulty lies with the signal itself. However, Garvey (1953b) found that Ss disagreed about the speech sounds that were rendered unintelligible by compression of the words in which they occurred. Garvey explained this finding in terms of the differential experience of Ss with respect to the words in question.

In this connection, Henry (1966) found words which occur with greater frequency in general language, as indicated by the Lorge-Thorndike count, to be more intelligible when compressed in time by the sampling method than less frequently occuring words.

Intelligibility and Anatomical Damage

The intelligibility of time compressed words is influenced by hearing capacity. In research reported by deQuiros (1964), Ss with normal hearing, Ss with hearing losses due to peripheral damage, and Ss with hearing losses due to central damage were given articulation tests with short sentences presented at 140, 250, and 350 wpm. For normal Ss, an increase in intensity of 10 db was required to reach threshold intensity (50% of the words identified correctly) with each increase in word rate. For Ss with hearing losses due to peripheral damage, the intensity required for threshold increased as word rate was increased in a similar manner, but each threshold intensity was somewhat higher than the corresponding intensity for Ss with normal hearing. For Ss with hearing losses due to central damage, the increase in intensity required for threshold as word rate was increased, was relatively large and positively accelerated.

Comprehension and Word Rate

There are several studies in which comprehension has been measured as a function of word rate. Within the range extending from 126 to 172 wpm, Diehl, White, and Burk (1959) found listening comprehension to be unaffected by changes in word rate. In the range extending from 125 to 225 wpm, Nelson (1948) and Harwood (1955) found a slight but insignificant loss in comprehension as word rate was increased. Fairbanks, Guttman, and Miron (1957) found little difference in the comprehension of listening selections presented at 141, 201, and 282 wpm. Thereafter, comprehension, as indicated by percent of test questions correctly answered, declined from 58% correct at 282 wpm to 26% at 470 wpm. Foulke, Amster, Nolan, and Bixler (1962), using both technical and literary listening selections, found comprehension to be only slightly affected by increasing word rates up to 275 wpm. However, in the range extending from 275 to 375 wpm, they found an accelerated decrease in comprehension as word rate was increased. The two studies last cited are in agreement regarding the finding that comprehension declined only slightly until a word rate in the neighborhood of 275 wpm was reached, but much more rapidly thereafter.

In each of the studies just mentioned, word rate was varied through a limited range. It is possible, by combining their results, to gain an impression of the relationship between word rate and listening comprehension over a wide range of word rate values. However, because these studies were conducted at different places and times, and with different Ss, listening selections and measuring instruments, the picture that emerges may not be completely dependable.

For this reason, an experiment was performed (Foulke, 1967) in which twelve groups of Ss, drawn at random from the same population, were tested for listening comprehension at twelve word rates, starting at 125 wpm and progressing in steps of 25 wpm through 400 wpm. After listening, each S completed a multiple choice test, covering the facts and implications of the listening selection.

A test score, corrected for guessing, was determined for each S, and these corrected scores were averaged for each experimental group to produce the means used in plotting Fig. 2. In this figure, the X-axis is scaled in terms of word rate, and the Y-axis is scaled in terms of mean percent of corrected comprehension scores. Although the curve shows some fluctuation, the relationship it suggests is one in which listening comprehension is little affected by increasing word rate in the range bounded by 125 and 250 wpm. Beyond this point, however, comprehension begins to decline at an accelerated rate. The significance of this relationship was examined by an analysis of variance. The significance of the differences among means was examined by the Newman-Keuls Test for Ordered Pairs of Means. The results of this analysis are shown in Table I. This is cast in matrix form with the word rates at which tests were conducted arranged down the left hand margin, and across the top of the matrix, in order of increasing magnitude. Entered in each row, under the appropriate column heading, are the word rates for which comprehension scores were not significantly different from the comprehension score associated with the word rate in the left hand margin that identifies that row. The results presented in Table I reinforce the impression suggested by Fig. 1. The pattern formed by the entries in this table clearly depict the nature of the relationship between word rate and listening comprehension. Listening comprehension does not vary significantly as word rate is increased in the range extending from 125 to 250 wpm. However, further increases in word rate result in losses in comprehension that are too large to be accounted for by chance fluctuation of test scores.

Comprehension and the Method of Compression

McLain (1962) and Foulke (1962), using Ss which were naive with respect to compressed speech, and unaccustomed to reading by listening, compared the comprehension of speech compressed by the sampling method with the comprehension of speech compressed by the speed changing method. In both instances, a slight but statistically significant difference was found in favor of the sampling method. However, in a similar experiment in which blind school children, who were accustomed to reading by listening, served as Ss, Foulke (1966a) found no statistically significant difference favoring either method.

The relationship between word rate and listening comprehension for speech compressed by the speed changing method, and by the sampling method, is shown in Fig. 3. In this figure, word rate,

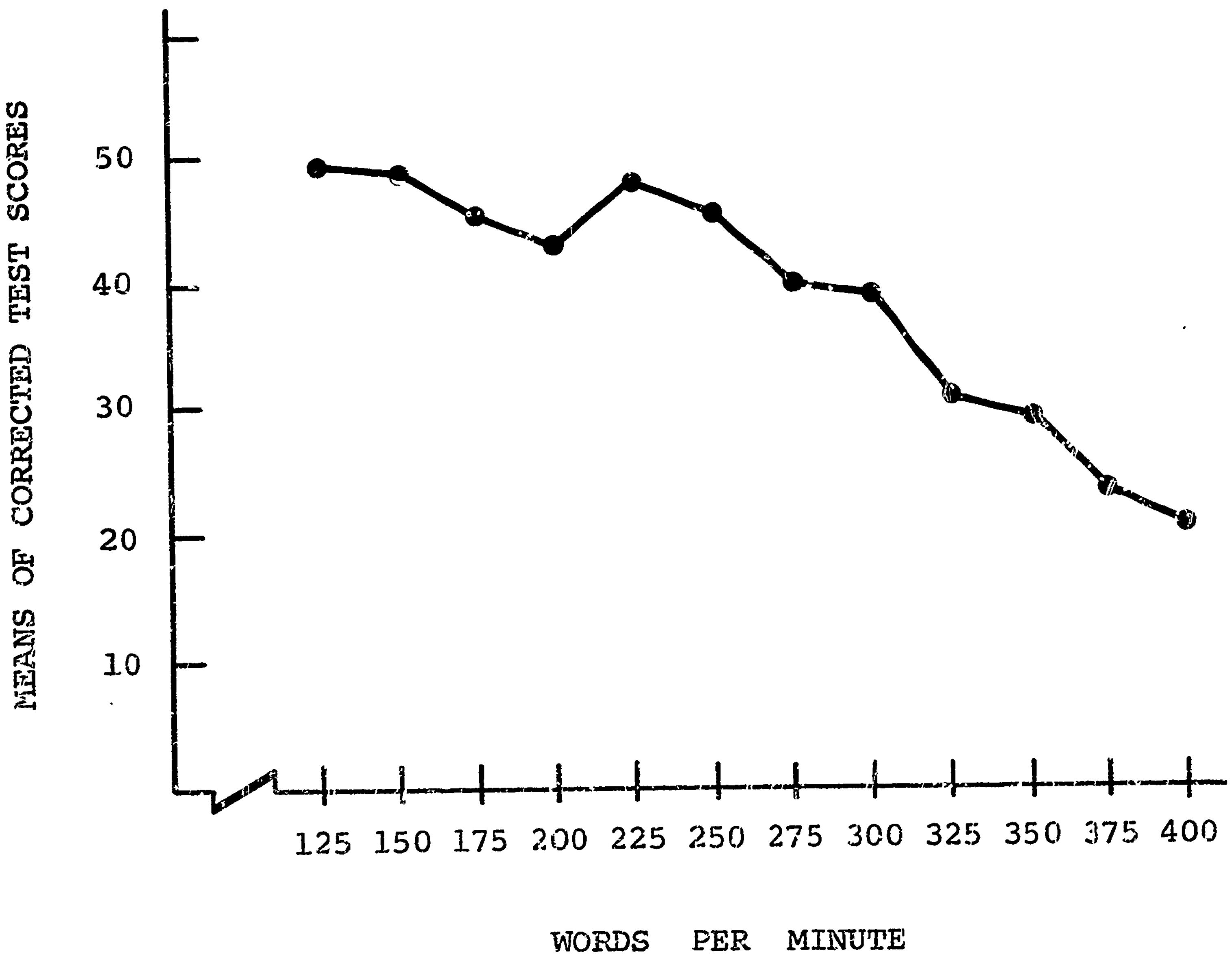

Fig. 2. Listening comprehension as a function of word rate.

TABLE I

NEWMAN-KEULS ANALYSIS OF THE SIGNIFICANCE OF

DIFFERENCES AMONG GROUP MEANS

WPM	125	150	175	200	225	250	275	300	325	350	375	400
125	125	150	175	200	225	250		300				
150	125	150	175	200	225	250	275	300				
175	125	150	175	200	225	250	275	300				
200	125	150	175	200	225	250	275	300				
225	125	150	175	200	225	250	275	300				
250	125	150	175	200	225	250	275	300				
275		150	175	200	225	250	275	300	325	350		
300	125	150	175	200	225	250	275	300	325	350		
325							275	300	325	350		
350							275	300	325	350	375	
375									325	350	375	400
400											375	400

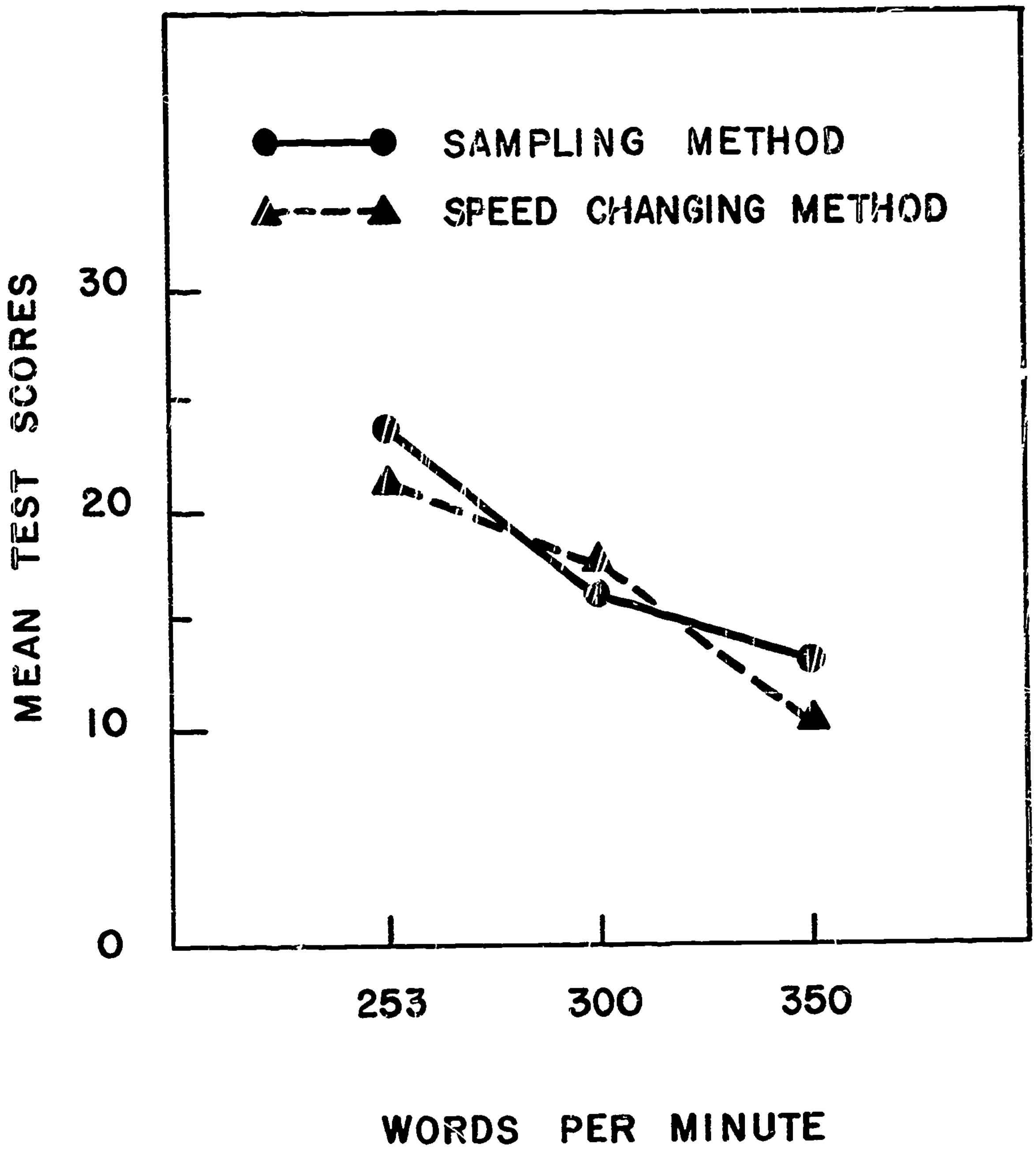

Fig. 3. Listening comprehension as a function of method of compression.

on the X-axis, is plotted against comprehension, on the Y-axis. The dotted curve describes results obtained with the speed changing method, and the solid curve describes results obtained with the electromechanical sampling method. The figure suggests that compression has approximately the same effect on comprehension, for both methods.

The conclusion suggested by the results of the experiments just cited is that the superiority of the sampling method is slight, and that this superiority may be erased by experience in reading by listening. It should be pointed out that this conclusion is only tentative. In the studies just cited, all Ss were naive with respect to speech compressed in time by the sampling method. The blind Ss probably had varying amounts of experience in listening to speech compressed in time by the speed changing method, since many blind people listen to Talking Book Records at a faster speed than the one used during recording. Furthermore, the tests were made in a limited range of word rates.

The Differential Effects of Compression on Word Intelligibility and Listening Comprehension

Experiments such as those reported by Garvey (1953a), Kurtzrock (1957), and Fairbanks and Kodman (1957), are in close agreement regarding the effect of compression by the sampling method on word intelligibility. When half or more of a word has been discarded, intelligibility is only moderately affected, and intelligibility is not completely lost when only 15% of a word remains. On the other hand, experiments such as those reported by Fairbanks, Guttman, and Miron (1957a), Foulke, et al. (1962), and Foulke (1967), generally agree in suggesting that comprehension begins to decline when only 45% of each of the words comprising a listening selection has been discarded. Taking these results together, it seems clear that listening comprehension is affected differently than word intelligibility by the time compression of speech by the sampling method. However, this conclusion is weakened somewhat by its dependence upon experiments that differed more or less with respect to stimulus materials, S populations, experimental plans, and techniques of measurement. A more satisfactory approach would be to demonstrate this differential effect in a single experiment. Consequently, an experiment was performed (Foulke & Sticht, 1967) in which listening material (PB words and a listening selection), compressed by 5 different amounts, was used in testing 5 experimental groups. Each S received an intelligibility score (the percent of correctly identified PB words) and a comprehension score (the percent of correctly answered multiple choice items in a test of listening comprehension). Mean intelligibility

scores and mean comprehension scores for the five experimental groups were used in plotting the curves in Fig. 4. In this figure, the five time compressions employed in this experiment are displayed along the X-axis. The entry below each compression value refers to the word rate that would result if connected discourse at a normal word rate of 175 wpm (Johnson, Darley, and Spriestersbach, 1963, pp. 202-203) were compressed by that amount. Percent correct for the two dependent variables is scaled on the Y-axis. As the amount of compression was increased, both intelligibility and comprehension decreased. However, comparison of the two curves reveals that intelligibility was always superior to comprehension and that intelligibility was affected much less than comprehension by increasing the amount of compression. The impression conveyed by Fig. 4 was confirmed by an analysis of variance of intelligibility and comprehension test scores. Both variables, and their interaction, were significant beyond the .001 level.

It was, of course, expected that comprehension scores would be lower than intelligibility scores. The demonstration of comprehension imposes a much more complex task on the listener than does the demonstration of intelligibility. It is the finding that the difference between intelligibility and comprehension scores increases as the amount of compression is increased that requires additional explanation.

One possibility is that the progressively larger loss in comprehension is a consequence of the cumulative effects of the relatively smaller losses in intelligibility. The data of the experiment were examined for this possibility in the following manner. All of the Ss tested at a given compression were separated into a high and a low scoring group, on the basis of their comprehension test scores. The difference between the means of the intelligibility scores of the two groups formed in this manner, was tested for significance. In all but one case, (the 59% compression group) the difference between means did not reach significance at the 5% level. This finding suggests that, with respect to the results of the present experiment, poor comprehension cannot be satisfactorily explained by low intelligibility for individual words. In any case, it is well known that it is not necessary for all of the units of a message to be intelligible in order for the message to be received accurately (Miller & Selfridge, 1950; Attneave, 1954). Because of prior learning, the listener is able to reconstruct the message on the basis of reduced cues. He makes use of the sequential dependencies in grammatical speech and the meaningfulness of the message in supplying missed words.

The results of Garvey (1953b), of Klumpp and Webster (1961), and of Henry (1966), reviewed earlier in this paper (see pp. 9-10), are pertinent in this regard. They found that increasing the number of phonemes in a word, or the number of words in a phrase, while holding constant the amount of compression in time, improved intelligibility. Increasing the number of phonemes in a brief message, should permit the listener to take advantage of his estimates of sequential dependency in reducing his uncertainty regarding elements of the message.

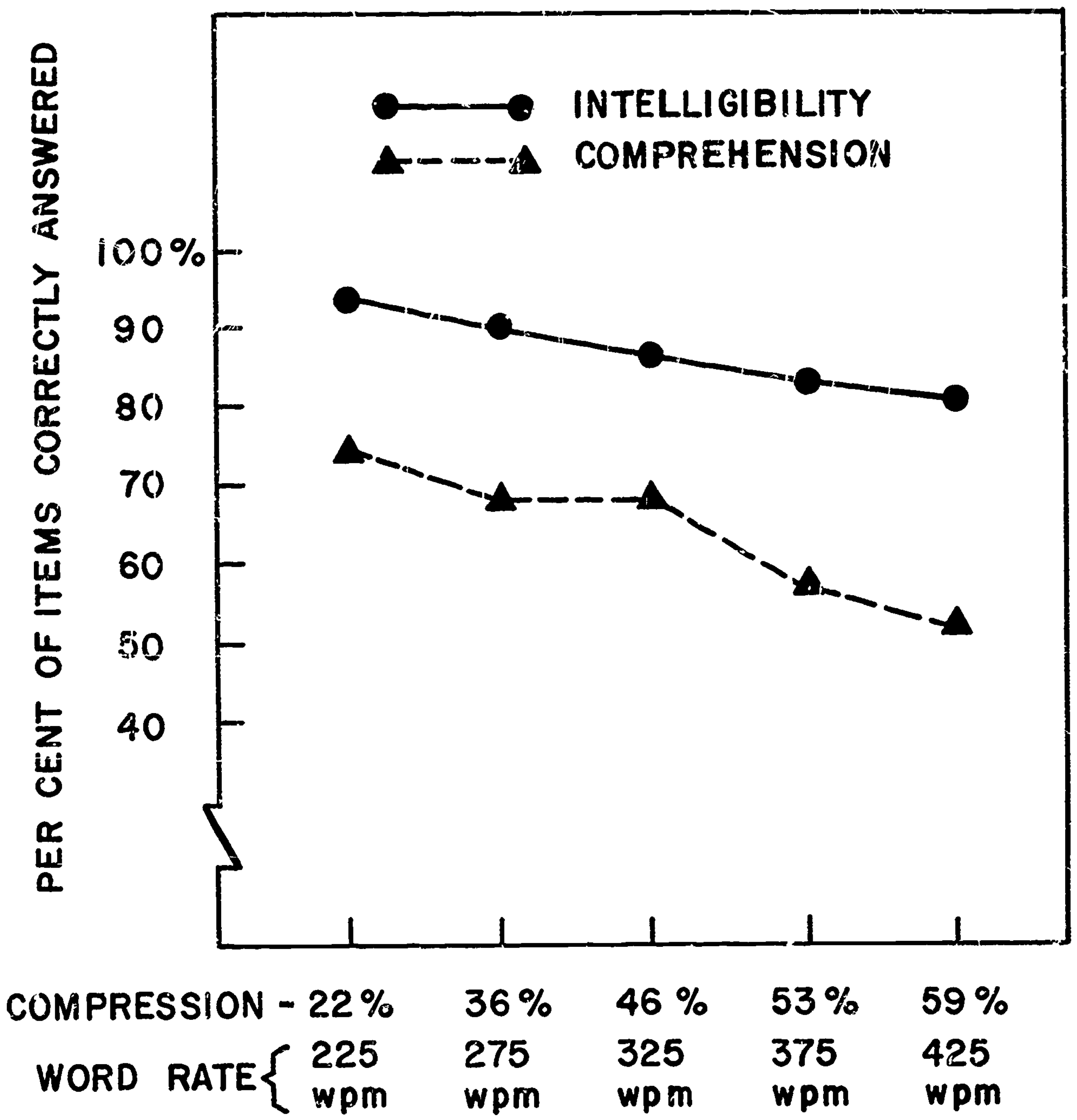

Fig. 4. Word intelligibility and listening comprehension as a function of percent of compression.

The net effect of the listener's ability to utilize cues of the sort just mentioned should be to free him from strict dependence upon the intelligibility of single words in comprehending messages composed of sequences of words.

It has already been shown that listening selections composed of words differing widely in intelligibility are approximately equal in comprehensibility (See Fig. 3). Of course, there is no doubt that if words are completely unintelligible, messages formed from them will be incomprehensible. However, one is tempted to draw the conclusion that once words have reached a certain level of intelligibility, further improvements in intelligibility will contribute little to the comprehensibility of messages formed from such words.

It is possible, by combining the speed changing method and the sampling method for the time compression of speech, to hold constant the rate at which speech sounds occur while varying the amount of distortion in vocal pitch. That is, if for each of several versions of a listening selection, the two methods for the time compression of speech are combined in different proportion to produce the same final accelerated word rate, the resulting versions of the listening selections will vary with respect to the amount of distortion in vocal pitch. Since there is a strong relationship between distortion in vocal pitch and word intelligibility, this scheme provides a way of varying word intelligibility systematically. Of course, the versions resulting from such a treatment will also vary with respect to the amount of speech information that has been discarded. But, as has already been shown, the sampling method has a relatively small influence on word intelligibility.

Taking advantage of this possibility, an experiment was performed (Foulke, 1967) in which the intelligibility of words in a listening selection was varied while the word rate was held constant. Five groups of Ss listened to five different versions of a listening selection, presented at a rate of 325 wpm. Proceeding from Version 1 to Version 5, there was a progressive elevation in vocal pitch. Following this listening experience, Ss completed a multiple choice test of the facts and implications of the listening selection.

Each S's score was the number of test items correctly answered. The means of these test scores, for the five experimental groups, are shown in Fig. 5. The vocal pitches employed in the experiment, arranged in order of increasing pitch, are shown on the X-axis. The Y-axis is scaled in terms of test scores. It is clear from this figure that differences in vocal pitch, and therefore presumably in word intelligibility, had practically no effect on listening comprehension. An analysis of variance of test scores indicated no significant difference in listening comprehension that could be associated with the experimental treatments. Therefore, it seems reasonable to conclude that within the range in which intelligibility was varied in this experiment, it had no influence on listening comprehension.

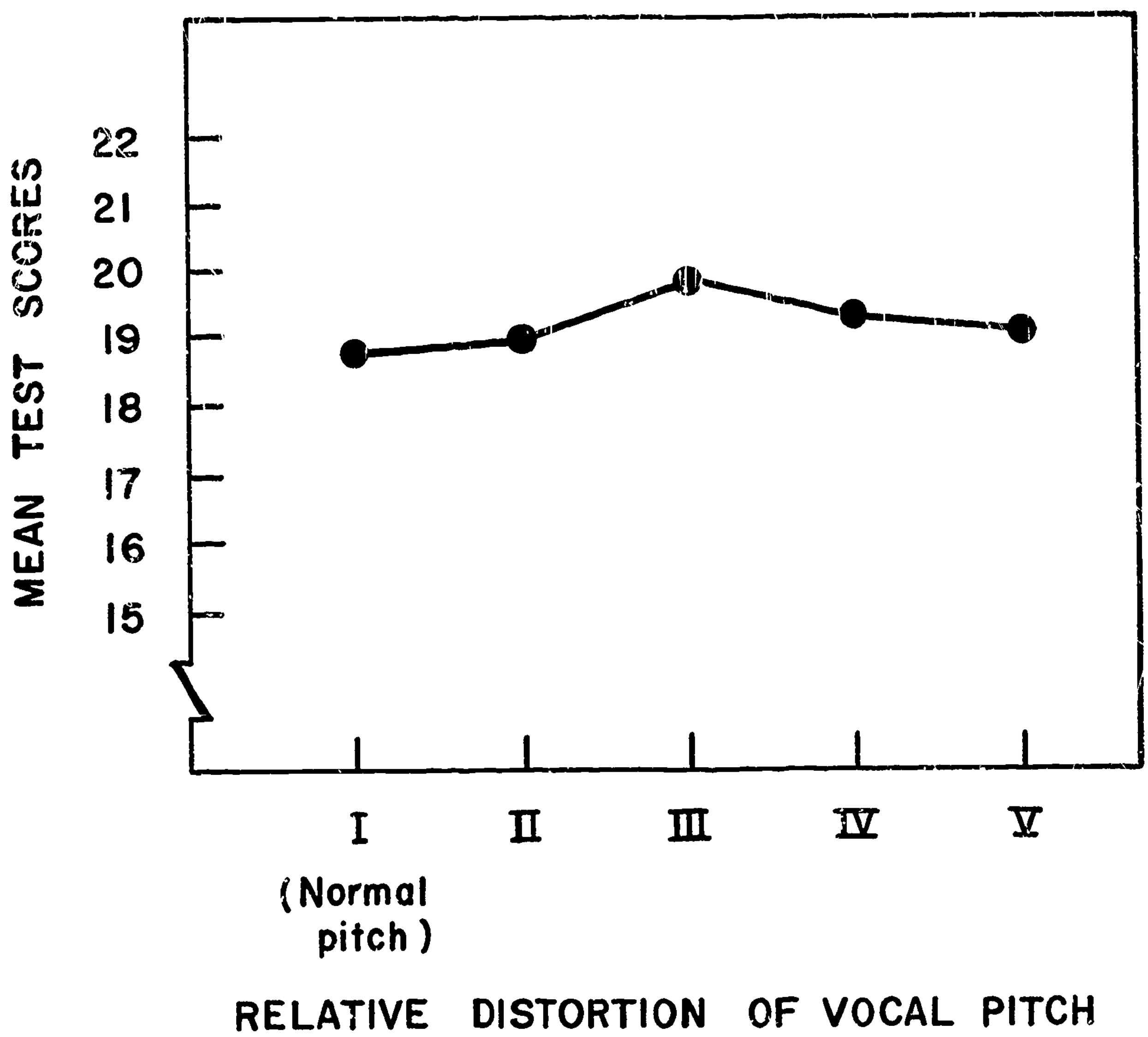

Fig. 5. Listening comprehension as a function of word intelligibility.

Evidence of another sort for the hypothesis that, within limits, variation in the intelligibility of single words has little effect on the comprehension of meaningful discourse composed of such words was provided by the results of another experiment (Foulke, Unpublished data). A professional oral reader produced three renditions of a listening selection: one at 149 wpm, one at 164.6 wpm, and one at 195.7 wpm. These renditions were then compressed, by the sampling method, to the same final word rate of 275 wpm. In order to achieve the same final word rate, it was necessary to compress the three selections respectively to 71%, 60%, and 50% of their original durations. It seems reasonable to assume that, in most cases, the durations of individual words were compressed by the same fraction as the durations of the renditions in which they occurred.

Three groups of Ss heard and were tested for comprehension of the three compressed renditions of the listening selection. Mean test scores for the three groups were practically the same, and an analysis of variance of test scores revealed no differences among means that could be associated with the experimental treatment provided.

Since the three renditions differed considerably with respect to the discarded fractions of words, there should be a systematic change in word intelligibility from rendition to rendition. This change would not be large since the compressions used in the experiment were not of sufficient magnitude to have a profound effect upon intelligibility. Nevertheless, the differences among the three renditions with respect to word intelligibility should have been at least measurable. To the extent that word intelligibility did vary from rendition to rendition, it is clear that it had no effect on listening comprehension.

Within the range in which intelligibility was varied in the two experiments just cited, it exerted no influence on the comprehension of connected speech. If intelligibility had been degraded sufficiently, there doubtless would have been a loss in comprehension. Nevertheless, within limits, comprehension does not appear to depend very heavily upon the intelligibility of single words. There is apparently enough redundancy in spoken language so that many words can be transmitted imperfectly or not at all without interfering seriously with listening comprehension. As a listener acquires experience with his language -- its grammar and its conventional forms -- he acquires information about the probabilities associated with the occurance of particular words, given the occurance of particular preceding words. Similarly, the context of meanings, aroused by a listening selection, reduces the listener's uncertainty, at any given instant, regarding the words and phrases that are to follow. The listener is able to use information about the occurance of words, phrases, and sentences in reconstructing imperfectly transmitted speech.

Comprehension and the Rate of Occurance of the Discriminable Elements of Speech

When the results of the experiments reviewed so far are considered collectively, a pattern of relationships begins to emerge. Word intelligibility, as defined by the measuring operations employed in its assessment, is seriously degraded when words are compressed by the speed changing method. When the sampling method is used, a compression of considerable magnitude is required before word intelligibility is seriously degraded (See Fig. 1). Listening comprehension, as defined by the measuring operations employed in its assessment, is only moderately affected by compression until a compression sufficient to produce word rates in the neighborhood of 275 wpm is surpassed (a compression of this magnitude would have little effect upon word intelligibility) but seriously affected thereafter (See Fig. 2). Thus, over a wide range, the relationship between compression and listening comprehension is not the same as the relationship between compression and word intelligibility (See Fig. 4).

The demonstration of comprehension imposes a much more complex task on the listener than does the demonstration of intelligibility. The behavior upon which the measurement of intelligibility depends implies registration of the stimulus word, some kind of short term memory storage, and the transduction of the stored item to an overt response. On the other hand, the behavior on which the measurement of comprehension is based implies continuous registration and short term storage of stimulus material, the continuous encoding, or simplification by reorganization, and selective discarding, of stimulus material so that it can be transferred to long term memory storage, and a final decoding step required for the transduction of material in long term storage to overt behavior. The model implied by these operations resembles models of the sort proposed by Sperling (1963), or Broadbent (1957). It seems reasonable to suppose that processing operations of the sort just mentioned would require time. With the additional assumption that the stored traces of stimulus events decay, so that after a short period of time they are no longer available for processing, a tentative explanation of the differential effects of compression on word intelligibility and listening comprehension is possible. In the situation in which the listener is tested for the intelligibility of words or short phrases or sentences, he receives stimulus input at a relatively slow rate. There is ample time to perform the processing operations required to transduce the stored effects of stimulation to an overt response. And, even though the memorial representation of the stimulus begins to decay immediately, it can be processed soon enough after registration so that its availability is not impaired. When, on the other hand, the listener must perform those processing operations needed to prepare a stimulus input, consisting of a meaningful

sequence of spoken words, for the subsequent demonstration of comprehension, he must process continuously a store to which new additions, in the form of the memorial representations of stimulus events, are being made continuously. If items are added to short term storage at a faster rate than the rate at which they can be processed, they will begin to accumulate. This accumulation will, however, be limited by the decay of items in store. When, because of a fast word rate, processing lags behind the addition of items to storage, some selection must be made among the accumulated items for processing. This selection may be made in terms of the relative availabilities of items. Those memorial representations of stimulus events that have been added to storage more recently will have had less time in which to decay, and will therefore be more available for processing. A consequence of this supposition is that when word rate does not exceed the processing rate, words will be processed in the order of their arrival. However, when word rate exceeds the processing rate, there will be a confusion of the order in which they are processed.

Many experiments in the general area of verbal learning and cognitive processing have produced results consistent with the point of view presented here. It has been shown repeatedly, in studies of verbal learning, (Miller, 1951, p. 212; Osgood, 1953, p. 505) that the difficulty of a learning task is increased by increasing the number of items in the list to be learned, and by decreasing the interstimulus interval. Of course, when continuous speech is compressed, the number of words per unit time is increased, and the intervals between words are decreased. On the other hand, compression cannot have this effect when the intelligibility of a single word is determined.

An experiment has been reported by Aaronson (1967), in which Ss were required to repeat sequences of heard digits. The acoustical durations of digits were adjusted by means of time compression so that the reproduction of each digit required 150 msec. Three different interstimulus intervals were represented in the experiment. In repeating sequences of heard digits, the number of errors, of both omission and order, made by Ss increased with decreasing duration of the interstimulus interval. Since the stimulus "on" time was the same for all spoken digits in all conditions of the experiment, this change in performance must have been the direct result of reducing the duration of the interstimulus interval, and hence the time available for perceptual processing. Insufficient processing time should, according to the model suggested here, result in errors of both omission and order.

Another variable in Aaronson's experiment has at least parenthetical relevance for the present discussion. It seems reasonable that stimuli which are, to some degree, unclear or illegible, should require more time for registration than legible stimuli. To test this hypothesis, stimuli were presented at three different signal to noise ratios. As expected, more errors were made in repeating sequences composed of stimuli presented with a poorer signal to noise

ratio. Using the speech compression devices presently available, there is some deterioration of signal quality when speech is compressed by the sampling method, and this deterioration increases slightly as the amount of compression is increased. This deterioration may be regarded as a degradation of the signal to noise ratio and hence, in addition to processing problems, there may be some registration problems that interfere with the perception of highly compressed speech.

In an experiment now in progress at the University of Louisville, Ss are required to repeat as many words as they can of sentences that are presented at an approximate rate of 500 wpm. When sentences are long, many words are missed, and the remaining words are often not repeated in the order in which they were presented. Again, as in the case of Aaronson's study, these are the kinds of errors that one would expect when the memorial representations of stimulus events accumulate at a faster rate than the rate at which they can be processed.

The two experiments just reported constitute examples in which discrete sequences of items have been used to demonstrate the functioning of the model. In describing the manner in which continuous speech is processed, it may be useful to introduce the concept of channel capacity (Miller, 1953 & 1956). According to this concept, a communication channel, in this case the listener, has a finite capacity for handling information. (This capacity would be defined by the rate at which the memorial representations of stimulus events could be processed for long term storage.) As the amount of information applied to the input of the channel is increased, there is a corresponding increase in the amount of information transmitted by the channel, until channel capacity is reached. (At this point, the input rate would match the processing rate.) Further increases in the input rate cannot be handled by the channel, with the result that some information is lost. Assuming normal speech to occur at a rate that is well below channel capacity, increasing word rate should have little effect on listening comprehension until channel capacity is reached. Beyond this point, further increases in word rate should result in progressively larger losses in listening comprehension. This is the situation that is described by the curve in Fig. 2 and, in general, by the results of other studies of the effect of word rate on listening comprehension (See p. 10).

The facts regarding silent visual reading rates may seem to constitute an obvious objection to the kind of explanation attempted here. Silent visual reading rates considerably in excess of 275 wpm, the word rate at which listening comprehension begins to deteriorate rapidly, are commonplace. However, because of the spatial display of information on the printed page, the reader is able to perform the perceptual operation referred to by Miller (1956) as "chunking".

in order to keep the input rate below his channel capacity, the visual reader reduces the number of elements with which he must contend by this "chunking" operation. With experience in reading, he learns to perceive not single words, but entire phrases or sentences. However, when language is displayed orally, it is displayed sequentially in a temporal dimension, and the listener cannot perform the "chunking" operation.

Though detailed neurological support for the point of view presented here is not possible, it is always tempting to suggest a neurological basis for a psychological model. The evidence presented by deQuiros, and reviewed earlier in this paper, (see p. 10) is interesting to consider in this regard. It will be recalled that when articulation testing was conducted with accelerated speech for Ss with hearing losses due to peripheral damage, an increase in word rate had approximately the same effect on the increase in intensity required for threshold intelligibility as for normal Ss. But at each word rate at which tests were conducted, threshold intensity for Ss with peripheral damage was somewhat higher than threshold intensity for normal Ss. The primary effect of peripheral damage should be a loss of sensitivity of the organs of hearing, and the results reported by deQuiros should be a reasonable consequence of such damage. The positively accelerated increase in the intensity required for threshold with increasing word rate, obtained from Ss with central damage, suggests that the increased rate at which words were presented constituted a more serious problem for these Ss than for Ss with normal hearing. Damage to the central nervous system would be of the sort that might be expected to interfere with perceptual or cognitive operations. It may therefore be that Ss in this category found the reproduction of sentences presented at accelerated word rates an especially difficult task because of a decreased ability to perform the processing operations required for the perception of such sentences.

To conclude, an effort has been made to account for the observed differences in the effect on word intelligibility and listening comprehension of increasing word rate, by appealing to a model for short term memory that bears a general resemblance to models of the sort suggested by Sperling (1963). It is recognized that the proposed model, with respect to its ability to account for the perception of time compressed speech, is only rudimentary in conception. Its parameters would have to be more carefully quantified before the predictions needed to test its adequacy could be made, and such quantification would require a great deal of research that has not yet been performed. Nevertheless, at its present stage of development, the model does seem to render a plausable account of the results of many of the experiments that have been reported.

# # # #

References

Aaronson, D. The temporal course of perception in an immediate recall task. Journal of Experimental Psychology, 1967, In Press.

Attneave, F. Some informational aspects of visual form perception. Psychological Review, 1954, 61, 183-193.

Broadbent, D. W. A mechanical model for human attention and immediate memory. Psychological Review, 1957, 64, 205-215.

Calearo, C. & Lazzaroni, A. Speech intelligibility in relation to the speed of the message. Laryngoscope, 1957, 67, 410-419.

Campanella, S. J., Signal Analysis of Speech Time-Compression Techniques. Proceedings of the Louisville Conference on Time Compressed Speech. Louisville: Univ. Louisville, 1967.

deQuiros, J. B. Accelerated speech audiometry, an examination of test results. Translations of the Beltone Institute for Hearing Research, #17, 1964.

Diehl, C. F., White, R. C., & Burk, K. Rate and communication. Speech Monographs, 1959, 26, 229-232.

Fairbanks, G., Everitte, W. L., & Jaeger, R. P. Method for time or frequency compression-expansion of speech. Transactions of the Institute of Radio Engineers Profession Group on Audio, Jan. & Feb. 1954, 7-12.

Fairbanks, G., Guttman, N., & Miron, M. S. Effects of time compression upon the comprehension of connected speech. Journal of Speech & Hearing Disorders, 1957, 22, 10-19.

Fairbanks, G. and Kodman, F., Jr. Word intelligibility as a function of time compression. Journal of the Acoustical Society of America, 1957, 29, 636-641.

Fletcher, H. Speech & Hearing, New York: D. Van Nostrand, Co., Inc., 1929, 293-294.

Foulke, E. A comparison of two methods of compressing speech. Symposium Paper, Southeastern Psychological Association, Louisville, March, 1962.

Foulke, E. The comprehension of rapid speech by the blind - Part III., Semi-annual Progress Report, Cooperative Research Project #2430, Office of Education. Washington, D. C.: U. S. Department of Health, Education, and Welfare, 1965.

Foulke, E. Comparison of comprehension of two forms of compressed speech. Exceptional Children, 1966, 169-173.

Foulke, E. The comprehension of rapid speech by the blind - Part III., Interim Progress Report, Cooperative Research Project #2430, Office of Education. Washington, D. C.: U. S. Department of Health, Education, and Welfare, 1967.

Foulke, E., Amster, C. H., Nolan, C. Y., & Bixler, R. H. The comprehension of rapid speech by the blind. Exceptional Children, 1962, 29, 134-141.

Foulke, E. and Sticht, T. G. The intelligibility and comprehension of accelerated speech. Proceedings of the Louisville Conference on Time Compressed Speech. Louisville: Univ. Louisville, 1967, 21-28.

Foulke, E. and Sticht, T. G. A review of research on time compressed speech. Proceedings of the Louisville Conference on Time Compressed Speech. Louisville: Univ. Louisville, 1967, 3-20.

Garvey, W. D. The intelligibility of speeded speech. Journal of Exceptional Psychology, 1953(a), 45, 102-108.

Garvey, W. D. The intelligiblity of abbreviated speech patterns. Quarterly Journal of Speech, 1953(b), 39, 296-306.

Harwood, K. A. Listenability and rate of presentation. Speech Monographs, 1955, 22, 57-59.

Henry, W. G., Jr. Recognition of time compressed speech as a function of word length and frequency of usage. (Unpub. doctoral dissertation, Indiana Univ.), 1966.

Johnson, W., Darley, F., & Spriestersbach, D. C. Diagnostic Methods in Speech Pathology. New York: Harper & Row, 1963, 202-203.

Klumpp, R. G. & Webster, J. C. Intelligibility of time-compressed speech. Journal of the Acoustical Society of America, 1961, 31, 265-267.

Kurtzrock, G. H. The effects of time and frequency distortion upon word intelligibility. Speech Monographs, 1957, 24, 94.

McLain, Julie. A comparison of two methods of producing rapid speech. International Journal for the Education of the Blind, 1962, 12, 40-43.

Miller, E. C. Effects on learning of variations in oral presentation. (Unpub. doctoral dissertation, Univ. of Denver), 1954.

Miller, G. A. Language and Communication, New York: McGraw-Hill Book Co., Inc., 1951.

Miller, G. A. What is information measurement. American Psychologist, 1953, 8, 3-11.

Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Miller, G. A. & Licklider, J. C. R. The intelligibility of interrupted speech, Journal of the Acoustical Society of America, 1950, 22, 167-173.

Miller, G. A. & Selfridge, J. A. Verbal context and the recall of meaningful material. American Journal of Psychology, 1950, 63, 176-185.

Nelson, H. E. The effect of variations of rates on the recall by radio listeners of straight newscasts. Speech Monographs, 1948, 15, 173-180.

Osgood, C. E. Method and theory in experimental psychology. New York: Oxford University Press, Inc., 1953.

Scott, R. J. Temporal effects in speech analysis and synthesis. (Unpub. doctoral dissertation, Univ. of Michigan), 1965.

Sperling, G. A model for visual memory tasks. Human Factors, (Feb.), 1963, 19-31.

Woodworth, R. S. & Schlosberg, H. Experimental Psychology, New York: Henry Holt & Co., Inc., 1954, 33.